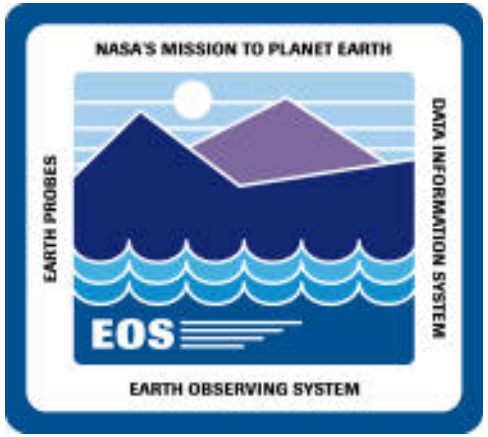


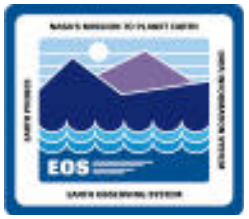
Detailed Ingest Subsystem Design

Carey Gire

cgire@eos.hitc.com

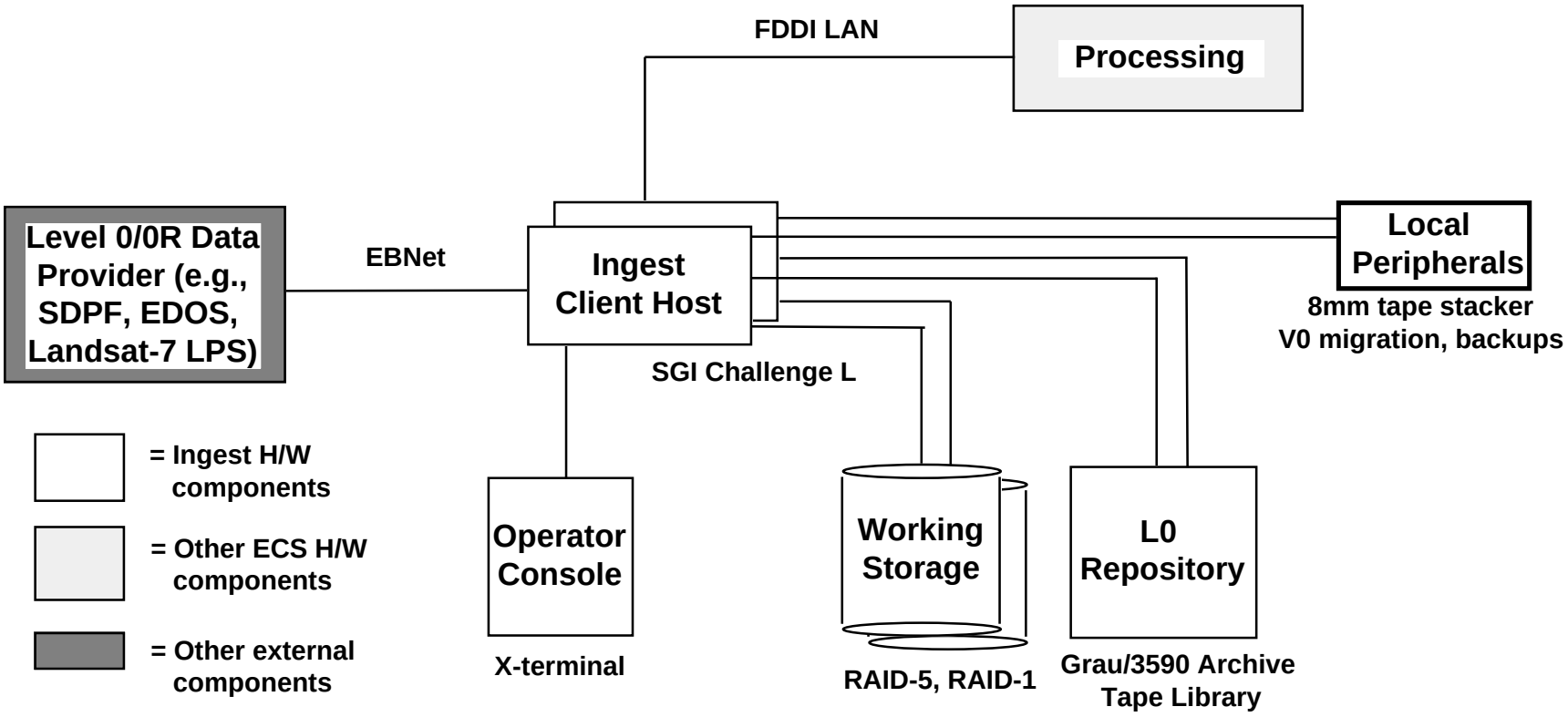
19 April 1996

Detailed Ingest Subsystem Design

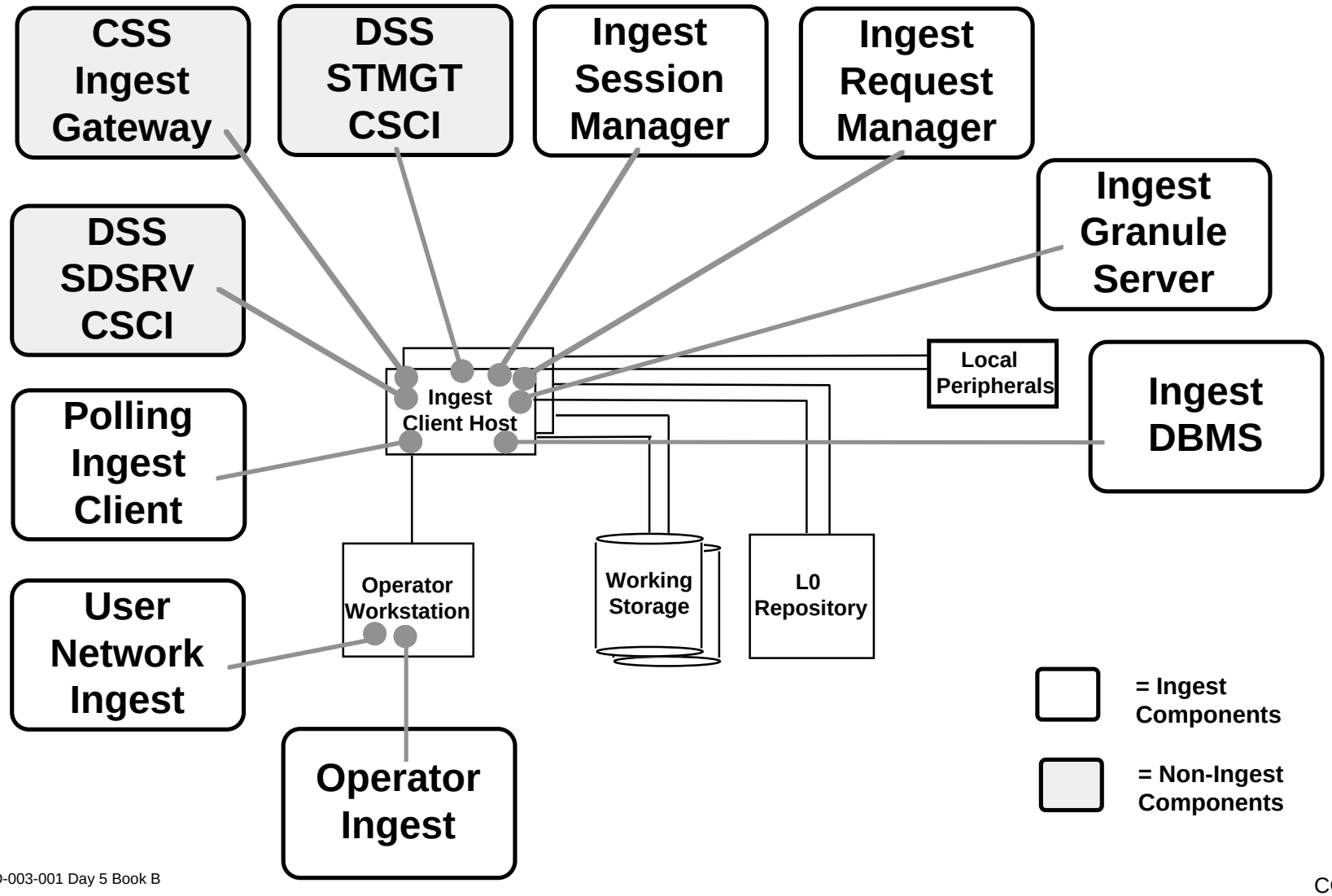


1. Ingest Subsystem Configuration
2. Release B Design Drivers and Constraints
3. Design Methodology
4. Equipment Allocation (e.g., LaRC)
5. Recoverability
6. Scalability

Ingest Client HWCI (ICLHW) Configuration



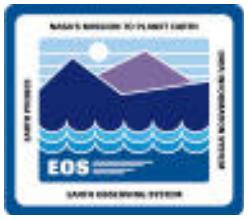
Process Mapping



Release B Hardware Design Drivers and Constraints

1. RMA Requirements
2. Daily Level 0 Ingest Volumes
3. Steady State Date Rates/Volume Accumulations
4. Transaction Patterns
5. Preprocessing Loads
6. Data Server Component Reuse

RMA Requirements



Reliability, Maintainability, Availability (RMA)

- 0.999 availability, 2 hours mean down time (MDT)
- 15 minute switchover to backup capability
- High-reliability storage of L0 data
- Maintenance of data integrity during system failures

ICLHW Impact

- Redundant processors, network connections
  - Cross-strapped access to storage devices
- MSS-controlled switchover to backup processor
- Robotics unit with redundant tape drive
- RAID-5 for working storage
- RAID-1 for Sybase transaction log

Daily Level 0 Ingest Volumes

Interface	Volume Per Day
GSFC - EDOS (MODIS L0)	70 GB
GSFC - ACRIM L0	<<1 GB
LaRC - Rel A L0 Sources	< 1 GB
LaRC - EDOS (CERES L0, MISR L0, MOPITT L0)	47 GB
LaRC - SAGE-III L0	< 1 GB
EDC - Landsat-7 L0R	139 GB*
JPL - RADAR ALT, ADEOS II L0	< 1 GB

*Ingested by the ICLHW HWCI, stored in DSS repository

Steady State Data Rates/Volume Accumulations

Site	I/O to and from Storage (MB/sec)	Working Storage Volume (GB)	Level 0 Repository Volume (GB)
GSFC	1.6	176	25,696
LaRC	1.2	119	17,374
EDC	3.2	174.25	0
JPL	0.006	0.675	98.6

Assumptions:

- Based on static analysis of the Technical Baseline
- Assume 1x retrievals from Working storage
- Working storage volume = 2.5 X daily L0 volume (1.25 for Landsat-7 L0R)
- Level 0 repository volume = 365 X daily L0 volume
- The Landsat-7 L0R data repository is supplied by Data Server

Ingest Request Transaction Patterns*

Site	Requests Per Day	Granules Per Day	Files Per Day	# DB Updates/ Accesses Per Day	Accumulation Per Day (KB)**
GSFC	800	2200	10000	200000	300
LaRC	100	200	800	40000	25
EDC	50	2000	14000	165000	150
JPL	50	600	6000	75000	50
NSIDC	10	20	200	30000	3
ORNL	7	10	100	30000	2

*Based on static analysis of Technical Baseline

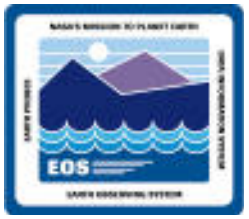
**Accumulation based on storage of summary history log information

Preprocessing Loads

Site	Metadata Parameters Extracted Per Day*	Conversions/ Reformatting Per Day
GSFC	45000	0
LaRC	4000	3**
EDC	40000	0
JPL	1500	0***
NSIDC	500	0
ORNL	<500	0

*Assuming 20 metadata parameters per data granule
**NMC ETA, FNL, MRL GRIB-to-HDF conversions (on Data Server Hardware)
***Potential AMSR conversions for SeaWinds

Reuse of Data Server Architecture



Shared Data Server Component Analyses

- **Component specifications**
- **Prototyping results**

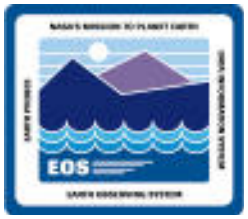
Similar Hardware Configurations

- **Minimizes hardware integration effort**
- **Minimizes long-term hardware maintenance**

Shared Peripheral Device Access

- **Allows load-balancing of pooled peripheral devices**
- **Common software interfaces, operations staff access and monitoring**

Hardware Design Methodology (Modeling/Analysis)



Evolutionary Process from Release A

Assumptions:

- Low Level 0 data retrieval rate after initial Processing access
- Expect most Processing accesses to come from Working Storage
  - Expect resultant low robotics transaction rate
- Large Instruments (MISR, MODIS) have Level 1A data; expect infrequent access of Level 0 data for reprocessing

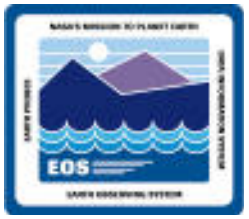
Steady State Modeling Used for Working Storage and Repository Designs

- Static modeling of Technical Baseline daily data volumes, rates
- Computation of data volume accumulations

Transaction Analysis Based on Technical Baseline

- Static modeling of robotics accesses
- Static modeling of Sybase accesses

Hardware Design Methodology—Sizing



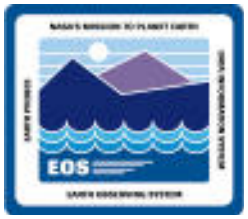
CPU Sizing

- I/O rates
- OODCE RAM overhead (based on ECS benchmarks of SGI OODCE)
- Preprocessing load
- Motif/Desktop RAM requirements

Disk Sizing

- Data Rates
- Data Residency
- AMASS Cache
- Ingest Data Base (managed by Sybase)
  - Data base accesses estimated based on transaction analysis
- Illustra Data Base
  - Data base accesses estimated based on daily granule insert rate
- Majority of RAID specified as SGI RAID 5 (SCSI-2 based)
  - < 1GB RAID-1 for DBMS transaction log

Hardware Design Methodology—Sizing (cont.)



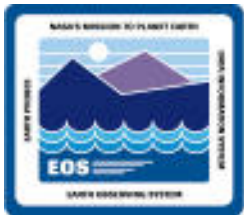
Robotics Sizing

- **Size estimated based on yearly accumulation of Level 0 data plus a small volume for accumulated Level 0 data without L1A counterpart**
- **Accesses estimated based on number of daily granule insert rate**
- **Tape recorder throughput = 3 X daily Level 0 input rate**

Ingest Subsystem LaRC Hardware

ICLHW Components	Equipment	Quantity
Ingest Host	2 CPU SGI Challenge L, 6 GB local disk, 256 MB RAM	2 ea.
Ingest Disk	119 GB of RAID disk, including an 11 GB AMASS cache	1 ea.
ATL Robotics	EMASS AML/E, single-arm robot, Quadro tower	1 ea.
Tape Recorders	3590 IBM NTP drives	2 ea.
Tape Media	NTP 10 GB cartridges	1750 ea.
Peripherals	8mm tape drive (with stacker)	2 ea.
Operator W/S	NCD HMX Pro X-terminal	1 ea.

Recoverability



Static Modeling Projection of System Recoverability

- **Processor switchover within 15 minutes (L3 Requirement)**
  - **During staffed operations only**
- **RAID-5 recovery of single disk failure with degraded service**
- **RAID-1 mirrored disk switchover for DBMS transaction logs**
- **Low duty cycle for robotics leads to low failure rate**
  - **Robotics arm may be easily fixed by on-site technician**
- **Redundant tape recorder supplied in the event of single tape recorder failure**

Ingest Client HWCI Scalability

- Ingest client hosts
  - Three backplane slots currently used
    * Two backplane slots available for additional CPUs, memory, SCSI-2 I/O boards
  - Currently 2 CPUs at GSFC, LaRC, JPL, and EDC
    * 2 more may be added to existing slot
    * Up to 4 CPUs (in multiples of 2) may be added at each backplane slot
  - Currently 256 MB of RAM at each site
    * Up to a total of 2 GB may be configured per memory board
- Working storage
  - Addition of RAID in 17 GB units
    * 85 GB may be added to the single existing I/O board
    * Up to 952 GB (7 slots per I/O board X 2 boards X 64 GB per RAID unit) may be added
- Level 0 data repository
  - Current AML/E robotics unit comes in one cabinet, holding up to 8 tape recorders
    * 1 additional cabinet (and 8 additional tape recorders) may be added
  - Each tape recorder requires a SCSI-2 slot; 2 currently used
    * Up to 14 additional may be added by adding 2 SCSI-2 I/O boards at the remaining backplane slots
- Ingest operator X-terminals/workstations
  - Additional X-terminals/workstations based on available FDDI slots